International Journal of Approximate Reasoning 52 (2011) 335–352

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

A belief function classifier based on information provided by noisy and dependent features

Paul-André Monney [a,*], Moses Chan [b], Paul Romberg [b]

[a] *Independent Consultant, 1975 East Pacific Street, Ely, IA 52227, USA*
[b] *Lockheed Martin Space Systems, Advanced Technology Center, Sunnyvale, CA 94089, USA*

ARTICLE INFO

Article history:
Received 14 February 2009
Received in revised form 25 November 2009
Accepted 26 November 2009
Available online 11 December 2009

Keywords:
Dempster–Shafer theory
Evidence theory
Object classification
Noisy measurements
Dependent features

ABSTRACT

A model and method are proposed for dealing with noisy and dependent features in classification problems. The knowledge base consists of uncertain logical rules forming a probabilistic argumentation system. Assumption-based reasoning is the inference mechanism that is used to derive information about the correct class of the object. Given a hypothesis regarding the correct class, the system provides a symbolic expression of the arguments for that hypothesis as a logical disjunctive normal form. These arguments turn into degrees of support for the hypothesis when numerical weights are assigned to them, thereby creating a support function on the set of possible classes. Since a support function is a belief function, the pignistic transformation is then applied to the support function and the object is placed into the class with maximal pignistic probability.

1. Introduction

Many papers present a solution to classification problems by using the theory of belief functions. In particular, in [1–3], the solution is presented using a case-based approach. This approach is compared to the model-based approach in [4] and pairwise classifier combination is presented in [5]. Another traditional methodology for classification problems is the Bayesian approach [6].

Ristic and Smets [7] present a method to classify objects into one of several possible classes $c_1, c_2, \ldots, c_m$, based on observed values of some features of the object. The starting point of the method is a list of continuous conditional density functions of a feature given each of the possible classes $c_i, i = 1, \ldots, m$. For each class c_i, the conditional density function is considered as being the pignistic density $Betf(x|c_i)$ of some unknown belief function on the real line [8–12]. Once the least committed belief function on $\mathbb{R}$ having $Betf(x|c_i)$ as pignistic density is determined [13,14], then, given an observed value x of the feature, the generalized Bayes theorem [13,15] is applied to derive a belief function Bel_x on the set of classes

$$C = \{c_1, \ldots, c_m\}. \tag{1}$$

The final step is to apply the pignistic transformation to Bel_x to find the pignistic probability $BetP(c_i|x)$ of each class c_i. The object is then placed in the class having the largest pignistic probability. In order to perform the classification, Ristic and Smets [7] assume that the exact value x of the feature is known and that the features are independent. As explained in [13], the generalized Bayes theorem can be extended to noisy feature values expressed by belief functions. Since a joint conditional density on the features representing dependencies among features can be transformed into a belief function by the least committed principle, this means that the method of Ristic and Smets [7] is in principle also capable of dealing with

* Corresponding author.
E-mail addresses: pmonney@monneyfusion.com (P.-A. Monney), moses.w.chan@lmco.com (M. Chan), paul.m.romberg@lmco.com (P. Romberg).

0888-613X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2009.11.008

imprecise and dependent features. In addition, Ref. [16] presents a method for constructing conditional belief functions based on sample data. In this paper, we present an alternative method to the Ristic and Smets method and its potential generalization to noisy and dependent features. Our method is capable of dealing with both noisy and dependent features.

Section 2 presents the model used for classification and the essential elements of probabilistic argumentation systems. It is shown how noisy and dependent features are represented in the model. Section 3 presents an application in the field of finance. Section 4 describes a common situation of feature dependence that presents itself when features are extracted from uncertain measurements of some characteristics of the object called signatures. The procedure used to compute the information about the features using kernel density estimation is explained and a numerical example is presented. Section 5 concludes the paper.

2. A model for noisy and dependent features

In this section we present a model for object classification that allows for noisy and dependent features.

2.1. Noisy features

The model consists of a knowledge base given by several logical rules. In order to describe this knowledge base, we first need to introduce a few key concepts and definitions. Let us denote by f a feature whose value can be any real number and let

$$\mathscr{I} = \{I_1, \ldots, I_n\} \tag{2}$$

denote a collection of real intervals forming a partition of $\mathbb{R}$, i.e.

$$I_i \cap I_j = \emptyset \text{ for } i \neq j, \quad \bigcup_{i=1}^{n} I_i = \mathbb{R}. \tag{3}$$

In a first step toward the creation of the rules to be placed in the knowledge base, these intervals are constructed in such a way that if we know for sure that feature f is within interval I_i, then the object belongs to a class within a certain subset C_i of C. In other words, if c denotes the correct unknown class of the object, we have the rule

$$f \in I_i \ \rightarrow \ (c \in C_i). \tag{4}$$

In addition, in a second step, it is assumed that there is some uncertainty attached to this rule. Specifically, knowing that $f \in I_i$ and that rule (4) is true does not always allow us to conclude that $c \in C_i$ with complete certainty. More precisely, if $f \in I_i$, then the certain conclusion $c \in C_i$ can be reached when some background condition is satisfied. For example, if the speed of an object is at least 525 km/h and less than 560 km/h, then we can conclude with complete certainty that it is either a bomber or a fighter when no other object type, not even any conceivable object type not in C, can have a speed in that range. Therefore, if v_i denotes the assumption that the background condition is satisfied, then we have the rule

$$v_i \ \rightarrow \ (f \in I_i \ \rightarrow \ (c \in C_i)). \tag{5}$$

On the other hand, if the background condition is not satisfied, i.e. the assumption v_i is false, then knowing that $f \in I_i$ does not allow us to conclude that $c \in C_i$. Our model then assumes that not only $c \in C_i$ cannot be concluded, but also nothing else can be concluded about the value of c, except that it belongs to the entire set C. Therefore, if nv_i denotes the assumption that the background condition is not satisfied, we have the rule

$$nv_i \ \rightarrow \ (f \in I_i \ \rightarrow \ (c \in C)). \tag{6}$$

Since $c \in C$ is always true we can write

$$(c \in C) = \top,$$

where $\top$ denotes the tautology. This shows that rule (6) now reads

$$nv_i \ \rightarrow \ (f \in I_i \ \rightarrow \ \top). \tag{7}$$

Using the equality sign to represent equivalent formulas, we have

$$\begin{aligned} nv_i \ \rightarrow \ (f \in I_i \ \rightarrow \ \top) &= nv_i \ \rightarrow \ ((\neg(f \in I_i)) \vee \top) \\ &= nv_i \ \rightarrow \ \top \\ &= \neg nv_i \vee \top \\ &= \top. \end{aligned} \tag{8}$$

If R_1 denotes rule (5) and R_2 denotes rule (6), then both R_1 and R_2 belong to the knowledge base. Therefore, the knowledge base is

$$KB = \{R_1, R_2, S_1, S_2, \ldots\}, \tag{9}$$

where $S_1, S_2, \ldots$ represent other rules besides R_1 and R_2. The knowledge base is represented by the formula

$$K = R_1 \wedge R_2 \wedge S_1 \wedge S_2 \wedge \cdots, \tag{10}$$

but since $R_2 = \top$ according to (8), we have

$$R_1 \wedge R_2 = R_1 \wedge \top = R_1 \tag{11}$$

and hence

$$K = R_1 \wedge S_1 \wedge S_2 \wedge \cdots. \tag{12}$$

This means that rule R_2 does not add any information to the knowledge base and can therefore safely be removed from it. Note that rule R_1, namely

$$v_i \rightarrow (f \in I_i \rightarrow (c \in C_i)), \tag{13}$$

can also be written as

$$v_i \wedge (f \in I_i) \rightarrow c \in C_i \tag{14}$$

because

$$\begin{aligned} v_i \rightarrow (f \in I_i \rightarrow (c \in C_i)) &= v_i \rightarrow ((\neg(f \in I_i)) \vee (c \in C_i)) \\ &= \neg v_i \vee \neg(f \in I_i) \vee (c \in C_i) \\ &= v_i \wedge (f \in I_i) \rightarrow c \in C_i. \end{aligned}$$

In general, it is unknown if the assumption v_i is true or false, but we assume that v_i is true with known probability $P(v_i)$. Formally, if a_i denote the variable with possible values v_i and nv_i, then a_i is called an assumption variable.

Our model also presumes that the exact value of feature f is not known precisely because there is some noise attached to its measurement, which is common in practical situations. Therefore, we consider that the information about the exact value of the feature is given by a continuous probability density function, which is denoted by $\varphi(x)$. This function allows us to compute the probability that feature f is within any interval I_i, namely

$$P(I_i) = \int_{I_i} \varphi(x)dx. \tag{15}$$

Then we can look at interval I_i as an assumption that is true with probability $P(I_i)$. Let a_f denote the assumption variable with possible values in the set

$$\mathscr{I} = \{I_1, \ldots, I_n\}. \tag{16}$$

In other words, the domain of the assumption variable a_f is $\mathscr{I}$ and

$$P(a_f = I_i) = P(I_i) = \int_{I_i} \varphi(x)dx. \tag{17}$$

Since $f \in I_i$ is only true with probability $P(I_i)$, we end up with rule R_1 being

$$(a_f = I_i) \wedge (a_i = v_i) \rightarrow (c \in C_i), \tag{18}$$

where both $P(a_f = I_i)$ and $P(a_i = v_i)$ are known.

To summarize, the knowledge base consists of the n rules

$$(a_f = I_i) \wedge (a_i = v_i) \rightarrow (c \in C_i), \quad i = 1, \ldots, n, \tag{19}$$

together with the probabilities $P(a_f = I_i)$ and $P(a_i = v_i)$. Of course, these probabilities satisfy

$$\begin{aligned} &\sum_{i=1}^{n} P(a_f = I_i) = 1, \\ &P(a_i = v_i) + P(a_i = nv_i) = 1. \end{aligned} \tag{20}$$

Such knowledge bases are called *probabilistic argumentation systems*, which we denote by PAS [10,17–20]. The roots of these systems have been laid by Laskey and Lehner [21] in the context of propositional logic. The essential idea of PAS is that belief functions can be derived by a sound combination of assumption-based reasoning and probability theory. In this context, degrees of belief come up as probabilities of *environments*, namely vectors of assumptions, allowing us to logically deduce the hypothesis. This perspective coincides with an interpretation of belief function provided by Pearl [22]. PAS has been further extended into a very general theory of *uncertain information* [23]. More recently, probabilistic assumption-based reasoning has been applied in the field of statistics by Kohlas and Monney [24], which expands from previous work of Monney on this topic [25]. Other relevant references are [26,27]. A description of a computer implementation of PAS called ABEL can be found in [28].

As an illustration, let us consider the situation described in Ristic and Smets [7]. The problem is the correct classification of unknown non-cooperative flying objects. There are three classes:

class c_1: commercial airplanes,
class c_2: large military aircraft, such as transport, bomber, and
class c_3: light and agile military aircraft (fighter planes).

The first classification feature, which we denote by f, is the speed of the object. The following table provides the minimum and maximum speeds (in km/h) for each class of objects.

By looking at these ranges, we can build the following six intervals (the last interval is really the union of two intervals, but this does not change the construction of the rules)

$$\begin{aligned} &I_1 = [400, 525), \quad I_4 = [725, 885), \\ &I_2 = [525, 560), \quad I_5 = [885, 950), \\ &I_3 = [560, 725), \quad I_6 = (-\infty, 400) \cup [950, +\infty). \end{aligned} \tag{21}$$

According to the speed ranges given in Table 1, we create the following six rules:

$$\begin{aligned} &(a_f = I_1) \wedge (a_1 = v_1) \rightarrow (c \in \{c_2\}), \\ &(a_f = I_2) \wedge (a_2 = v_2) \rightarrow (c \in \{c_2, c_3\}), \\ &(a_f = I_3) \wedge (a_3 = v_3) \rightarrow (c \in \{c_1, c_2, c_3\})), \\ &(a_f = I_4) \wedge (a_4 = v_4) \rightarrow (c \in \{c_1, c_3\}), \\ &(a_f = I_5) \wedge (a_5 = v_5) \rightarrow (c \in \{c_3\}), \\ &(a_f = I_6) \wedge (a_6 = v_6) \rightarrow (c \in \{c_1, c_2, c_3\}) \end{aligned}$$

and we take

$$P(a_i = v_i) = 0.9 \tag{22}$$

for all $i = 1, \ldots, 6$. The information about the value of the speed feature f is given by a probability density function $\varphi(x)$ which assigns the following probabilities to the six intervals:

$$\begin{aligned} &P(a_f = I_1) = 0.1, \quad P(a_f = I_4) = 0.2, \\ &P(a_f = I_2) = 0.2, \quad P(a_f = I_5) = 0.1, \\ &P(a_f = I_3) = 0.3, \quad P(a_f = I_6) = 0.1. \end{aligned} \tag{23}$$

The second feature considered by Ristic and Smets [7], which we denote by g, is the acceleration of the object. Table 2 provides the minimum and maximum acceleration (in g, the acceleration due to gravity) for each class of objects.

These ranges induce the intervals

$$\begin{aligned} &K_1 = [-7, -4), \quad K_4 = [1, 4), \\ &K_2 = [-4, -1], \quad K_5 = [4, 7), \\ &K_3 = [-1, 1), \quad K_6 = (-\infty, -7) \cup [7, +\infty). \end{aligned}$$

Let a_g denote the assumption variable associated with these intervals for feature g. Furthermore, let $\{u_i, nu_i\}$ denote the domain of the assumption variable b_i associated with the rule for interval K_i. Then, according to Table 2, we create the rules

$$\begin{aligned} &(a_g = K_1) \wedge (b_1 = u_1) \rightarrow (c \in \{c_3\}), \\ &(a_g = K_2) \wedge (b_2 = u_2) \rightarrow (c \in \{c_2, c_3\}), \\ &(a_g = K_3) \wedge (b_3 = u_3) \rightarrow (c \in \{c_1, c_2, c_3\}), \\ &(a_g = K_4) \wedge (b_4 = u_4) \rightarrow (c \in \{c_2, c_3\}), \\ &(a_g = K_5) \wedge (b_5 = u_5) \rightarrow (c \in \{c_3\}), \\ &(a_g = K_6) \wedge (b_6 = u_6) \rightarrow (c \in \{c_1, c_2, c_3\}) \end{aligned}$$

and we take

$$P(b_i = u_i) = 0.9 \tag{24}$$

Table 1
Speed range for each class of objects.

Class	Min	Max
Commercial (c_1)	560	885
Bomber (c_2)	400	725
Fighter (c_3)	525	950

Table 2
Acceleration range for each class of objects.

Class	Min	Max
Commercial (c_1)	−1	1
Bomber (c_2)	−4	4
Fighter (c_3)	−7	7

for all $i = 1, \ldots, 6$. The information about the value of the acceleration feature g is given by a probability density function $\tau(x)$ which assigns the following probabilities to the intervals:

$$\begin{aligned}&P(a_g = K_1) = 0.4, \quad P(a_g = K_4) = 0.1,\\ &P(a_g = K_2) = 0.1, \quad P(a_g = K_5) = 0.2,\\ &P(a_g = K_3) = 0.1, \quad P(a_g = K_6) = 0.1.\end{aligned} \tag{25}$$

The 12 rules, along with the corresponding assumption probabilities, form the knowledge base used for the object's classification. In general, given a knowledge base built from several features, the inference principles of probabilistic assumption-based reasoning can be applied to derive a belief function on the set C of possible classes for the object. By definition, for a hypothesis $H \subseteq C$, the quasi-support of H is the set of assumption vectors that allow us to logically prove H or make the knowledge base contradictory. A detailed presentation of probabilistic assumption-based reasoning and PAS can be found in Kohlas and Monney [10] and Haenni et al. [17]. Let $QS(H)$ denote the quasi-support of H and let $QS(\perp)$ denote the quasi-support of the contradiction, i.e. the set of assumption vectors that make the knowledge base contradictory. By definition, the degree of quasi-support of H is

$$qs(H) = P(QS(H)) \tag{26}$$

and the degree of contradiction is

$$qs(\perp) = P(QS(\perp)), \tag{27}$$

where P is the joint probability measure over all assumption variables in the knowledge base. Then the degree of support of the hypothesis H provided by the knowledge base is defined as the quantity

$$sp(H) = \frac{qs(H) - qs(\perp)}{1 - qs(\perp)}. \tag{28}$$

For now, let us assume that the assumption variables are independent, which means that P is the probability measure obtained by multiplying the individual assumption variables' probability distributions. In particular, we assume that the two features f and g are independent, which means that the assumption variables a_f and a_g are independent. The function

$$sp: \ 2^C \ \rightarrow \ [0,1], \tag{29}$$

defined in Eq. (28) is a belief function in the sense of Shafer [9]. However, this belief function does not reflect one's subjective beliefs about the correct class of the object, but rather represents the level of support for hypotheses provided by the knowledge base. The degree of support of H measures the level of reliability of the arguments for the hypothesis H. Once the support function sp has been computed, the pignistic transformation is applied to it to determine the pignistic probability $BetP(c_i)$ of each class c_i in C. The object is then placed in the class having largest pignistic probability.

In our example, using the PAS software ABEL,[1] we find that the support function sp on C is given by

$$\begin{aligned}&sp(\emptyset) = 0, \quad sp(c_1, c_2) = 0.044,\\ &sp(c_1) = 0, \quad sp(c_1, c_3) = 0.647,\\ &sp(c_2) = 0.044, \quad sp(c_2, c_3) = 0.812,\\ &sp(c_3) = 0.594, \quad sp(c_1, c_2, c_3) = 1.\end{aligned} \tag{30}$$

Note that we write $sp(c_1) = 0$ instead of $sp(\{c_1\}) = 0$ to keep the notation simple. This convention is used throughout the entire paper. The basic probability mass m associated with sp is

$$\begin{aligned}&m(\emptyset) = 0, \quad m(c_1, c_2) = 0,\\ &m(c_1) = 0, \quad m(c_1, c_3) = 0.053,\\ &m(c_2) = 0.044, \quad m(c_2, c_3) = 0.174,\\ &m(c_3) = 0.594, \quad m(c_1, c_2, c_3) = 0.135\end{aligned} \tag{31}$$

and the corresponding pignistic probabilities are

[1] The software ABEL can be downloaded for free from the Institute of Informatics of the University of Fribourg, Switzerland, http://diuf.unifr.ch/tcs/abel/.

$$BetP(c_1) = 0.0715, \quad BetP(c_2) = 0.176, \quad BetP(c_3) = 0.7525. \tag{32}$$

Therefore, given the information provided by the two features *f* and g, we conclude that the object belongs to class c_3, i.e. the object is a military fighter jet.

In order to explain in detail the calculations needed to find the support function *sp*, let us explicitly compute the degree of support for the hypothesis

$$H = \{c_1, c_3\}. \tag{33}$$

According to Eq. (28), we need to find the quasi-support of *H* and its probability $P(QS(H))$. Using ABEL, we find that the quasi-support of *H* is

$$QS(c_1, c_3) = (v_4 \wedge I_4) \vee (v_5 \wedge I_5) \vee (u_1 \wedge K_1) \vee (u_5 \wedge K_5) \tag{34}$$

and if we define

$$A = (v_4 \wedge I_4) \vee (v_5 \wedge I_5), \quad B = (u_1 \wedge K_1) \vee (u_5 \wedge K_5) \tag{35}$$

then *A* and *B* are probabilistically independent and we have

$$\begin{aligned} P(QS(H)) &= P(A \vee B) \\ &= 1 - P(\neg A \wedge \neg B) \\ &= 1 - ((1 - P(A))(1 - P(B))) \\ &= P(A) + P(B) - P(A)P(B). \end{aligned} \tag{36}$$

But

$$\begin{aligned} P(A) &= P(v_4 \wedge I_4) + P(v_5 \wedge I_5) - P(v_4 \wedge I_4 \wedge v_5 \wedge I_5) \\ &= P(v_4 \wedge I_4) + P(v_5 \wedge I_5) \end{aligned} \tag{37}$$

because

$$P(V_4 \wedge I_4 \wedge v_5 \wedge I_5) = 0 \tag{38}$$

since $I_4 \wedge I_5 = \perp$ because I_4 and I_5 are disjoint. Similarly, we obtain

$$P(B) = P(u_1 \wedge K_1) + P(u_5 \wedge K_5). \tag{39}$$

This implies that

$$\begin{aligned} P(A)P(B) &= P(v_4 \wedge I_4)P(u_1 \wedge K_1) + P(v_4 \wedge I_4)P(u_5 \wedge K_5) \\ &\quad + P(v_5 \wedge I_5)P(u_1 \wedge K_1) + P(v_5 \wedge I_5)P(u_5 \wedge K_5) \qquad (40) \\ &= P(v_4 \wedge I_4 \wedge u_1 \wedge K_1) + P(v_4 \wedge I_4 \wedge u_5 \wedge K_5) \\ &\quad + P(v_5 \wedge I_5 \wedge u_1 \wedge K_1) + P(v_5 \wedge I_5 \wedge u_5 \wedge K_5). \qquad (41) \end{aligned}$$

Then, according to (36), we finally obtain

$$\begin{aligned} P(QS(H)) &= P(v_4 \wedge I_4) + P(v_5 \wedge I_5) + P(u_1 \wedge K_1) + P(u_5 \wedge K_5) \\ &\quad - P(v_4 \wedge I_4 \wedge u_1 \wedge K_1) - P(v_4 \wedge I_4 \wedge u_5 \wedge K_5) \\ &\quad - P(v_5 \wedge I_5 \wedge u_1 \wedge K_1) - P(v_5 \wedge I_5 \wedge u_5 \wedge K_5). \end{aligned} \tag{42}$$

Let us denote by p_i the probability that assumption variable a_f takes on the value I_i, i.e.

$$p_i = P(a_f = I_i) \tag{43}$$

and by q_j the probability that assumption variable a_g takes on the value K_j, i.e.

$$q_j = P(a_g = K_j). \tag{44}$$

Furthermore, let us denote by r_i the probability that assumption variable a_i takes on value v_i, i.e.

$$r_i = P(a_i = v_i), \tag{45}$$

and by s_j the probability that assumption variable b_j takes on the value u_j, i.e

$$s_j = P(b_j = u_i). \tag{46}$$

Then, since the assumption variables are independent, Eq. (42) implies that

$$P(QS(H)) = r_4 p_4 + r_5 p_5 + s_1 q_1 + s_5 q_5 - r_4 s_1 p_4 q_1 - r_4 s_5 p_4 q_5 - r_5 s_1 p_5 q_1 - r_5 s_5 p_5 q_5. \tag{47}$$

Using the numerical values given in (22)–(25), we obtain

$$P(QS(H)) = 0.6642. \tag{48}$$

Regarding the quasi-support of the contradiction, we have

$$QS(\perp) = (v_1 \wedge I_1 \wedge K_1 \wedge u_1) \vee (v_1 \wedge I_1 \wedge K_5 \wedge u_5) \tag{49}$$

and hence

$$P(QS(\perp)) = P(v_1 \wedge I_1 \wedge K_1 \wedge u_1) + P(v_1 \wedge I_1 \wedge K_5 \wedge u_5) \tag{50}$$

since K_1 and K_5 are disjoint. Therefore, the degree of contradiction in the knowledge base is

$$P(QS(\perp)) = r_1 p_1 q_1 s_1 + r_1 p_1 q_5 s_5 \tag{51}$$

and using the numerical values given above we find

$$P(QS(\perp)) = 0.0486. \tag{52}$$

Finally, Eq. (28) implies that the degree of support of the hypothesis $H = \{c_1, c_3\}$ is

$$sp(H) = \frac{0.6642 - 0.0486}{1 - 0.0486} = 0.647, \tag{53}$$

which is the value given in Eq. (30).

2.2. Dependent features

Let us now see what happens when the features f and g are not independent. In this case, the information about these features is represented by a joint probability density, which we denote by $\psi(x,y)$. This density is not equal to the product of the marginal densities of f and g. As a consequence, the assumption variables a_f and a_g are also not independent and

$$P(I_i \wedge K_j) = \int_{I_i} \int_{K_j} \psi(x,y)\,dx\,dy. \tag{54}$$

Note that Eqs. (42) and (50) are still valid in case of dependence. Let $\mathscr{A}$ denote the set of assumption variables that are different from a_f and a_g. If we still assume that the variables in $\mathscr{A}$ are independent and assume that the variables in $\mathscr{A} \cup \{(a_f, a_g)\}$ are independent, which is reasonable, then we obtain

$$\begin{aligned} P(QS(H)) = {} & P(v_4)P(I_4) + P(v_5)P(I_5) + P(u_1)P(K_1) + P(u_5)P(K_5) \\ & - P(v_4)P(u_1)P(I_4 \wedge K_1) - P(v_4)P(u_5)P(I_4 \wedge K_5) \\ & - P(v_5)P(u_1)P(I_5 \wedge K_1) - P(v_5)P(u_5)P(I_5 \wedge K_5) \end{aligned} \tag{55}$$

and

$$P(QS(\perp)) = P(v_1)P(u_1)P(I_1 \wedge K_1) + P(v_1)P(u_5)P(I_1 \wedge K_5). \tag{56}$$

Using the notation introduced above, we get

$$P(QS(H)) = r_4 P(I_4) + v_5 P(I_5) + s_1 P(K_1) + s_5 P(K_5) - r_4 s_1 P(I_4 \wedge K_1) - r_4 s_5 P(I_4 \wedge K_5) - r_5 s_1 P(I_5 \wedge K_1) - r_5 s_5 P(I_5 \wedge K_5) \tag{57}$$

and

$$P(QS(\perp)) = r_1 s_1 P(I_1 \wedge K_1) + r_1 s_5 P(I_1 \wedge K_5). \tag{58}$$

In order to find the probability of I_i and the probability of K_j, we need to determine the information about feature f and feature g individually. This is done by marginalizing the joint feature density $\psi(x,y)$ to the variables f and g. In other words, the information about feature f is the density

$$\psi_f(x) = \int_{-\infty}^{+\infty} \psi(x,y)\,dy \tag{59}$$

and the information about feature g is the density

$$\psi_g(y) = \int_{-\infty}^{+\infty} \psi(x,y)\,dx. \tag{60}$$

Hypothetically, suppose that the information about the two features f and g is given by a bivariate normal distribution with parameters

$$\mu_f = 885, \quad \sigma_f = 95, \quad \mu_g = 3, \quad \sigma_g = 1.5, \quad \rho = 0.9, \tag{61}$$

where ρ is the correlation coefficient between the two features. Of course, in this case, the marginal of f is $N(885, 95)$ and the marginal of g is $N(3, 1.5)$. Then we obtain the following probabilities that are needed to compute the degree of support of H in this case of dependence

$$\begin{aligned}
&P(I_4) = 0.453, \quad P(I_4 \wedge K_1) = 0,\\
&P(I_5) = 0.253, \quad P(I_4 \wedge K_5) = 0.005,\\
&P(K_1) = 0, \quad P(I_5 \wedge K_1) = 0,\\
&P(K_5) = 0.248, \quad P(I_5 \wedge K_5) = 0.054,\\
&P(I_1 \wedge K_1) = 0, \quad P(I_1 \wedge K_5) = 0.
\end{aligned} \tag{62}$$

If we still take $r_i = 0.9$ for $i = 1, \ldots, 6$ and $s_j = 0.9$ for $j = 1, \ldots, 6$, then we obtain

$$P(QS(H)) = 0.811, \quad P(QS(\perp)) = 0 \tag{63}$$

and hence the degree of support of the hypothesis $H = \{c_1, c_3\}$ is

$$sp(H) = 0.811. \tag{64}$$

By comparison, if we assume that the two features are independent, i.e we take $\rho = 0$, then we obtain $P(QS(H)) = 0.717$. Since we still have $P(QS(\perp)) = 0$, we conclude that the degree of support for H in case of independence is

$$sp(H) = 0.717. \tag{65}$$

This technique to deal with situations of feature dependence can be applied to any probabilistic argumentation system. In general, the probability of the disjunctive normal form (DNF) representing the quasi-support of a hypothesis can be computed by the classical inclusion-exclusion formula from probability theory [10]. The absolute value of each term in the sum is the probability of a conjunction, which can be computed from the joint distribution on the assumption variables. This joint distribution reflects the dependence and independence relations that exist among those variables. Other methods for computing the probability of a DNF have been proposed in the literature [10,29]. It is important to note that the availability of the quasi-support of a hypothesis as an explicit disjunctive normal form, as in Eqs. (34) and (49), is essential to determine correct degrees of support in case of dependence. The ability to compute such DNF is a unique feature of software implementations of PAS, such as ABEL. Ref. [30] presents a model for the combination of dependent belief functions defined on finite frames of discernment.

3. Stock price analysis

In this section another application of the ideas discussed above is presented. When considering the price of a stock over time, we are interested in signs suggesting it might be time to buy or sell the stock. In the financial literature, these signs are called technicals and are supposed to help traders make better buy and sell decisions. A few decades ago, a technical called stochastic oscillator was proposed for this purpose. In order to define this technical, let us first introduce some notation. On a given trading day represented by a time index j, let

$$X_j^{close}, \quad X_j^{high}, \quad X_j^{low}$$

denote the stock price at the end of the day, the highest stock price during the day and the lowest stock price during the day. The stochastic oscillator is computed from stock prices over a period of N trading days. More specifically, if t denotes the time index of the present trading day and

$$N_t = \{t, t-1, \ldots, t-N+1\}$$

denotes the time indexes for the past N trading days, then the stochastic oscillator at time t is defined as

$$SO^t = 100 \cdot \frac{X_t^{close} - \min\{X_j^{low}, j \in N_t)}{\max\{X_j^{high}, j \in N_t) - \min\{X_j^{low}, j \in N_t\}}.$$

In this formula, the numerator is the difference between the present day's closing stock price and the lowest stock price over the last N trading days, whereas the denominator is the difference between the highest stock price and the lowest stock price over the last N trading days.

In addition to this standard stock technical, we introduce a second technical called random oscillator, which is quite similar to the stochastic oscillator. The random oscillator at time t is defined as

$$RO^t = 100 \cdot \frac{X_t^{close} - \min\{X_j^{close}, j \in N_t\}}{\max\{X_j^{close}, j \in N_t) - \min\{X_j^{close}, j \in N_t\}}.$$

In this formula, the numerator is the difference between the present day's closing stock price and the lowest closing price over the last N trading days, whereas the denominator is the difference between the highest closing stock price and the lowest closing stock price over the last N trading days. It is important to note that both SO^t and RO^t can vary between 0 and 100.

Based on these two oscillators, we are now in a position to introduce the so-called compound stochastic oscillator at time t, which is defined as the random variable CSO^t having a normal distribution with mean m_{SO^t} and standard deviation s_{SO^t}, where m_{SO^t}, s_{SO^t} are the mean and standard deviation of the data set

$$SO^k, \quad k = t, \ldots, t - N + 1.$$

In other words, we have

$$CSO^t \sim N(m_{SO^t}, s_{SO^t}). \tag{66}$$

Similarly, we define the compound random oscillator at time t as the random variable CRO^t having a normal distribution with mean m_{RO^t} and standard deviation s_{RO^t}, where m_{RO^t}, s_{RO^t} are the mean and standard deviation of the data set

$$RO^k, \quad k = t, \ldots, t - N + 1.$$

In other words, we have

$$CRO^t \sim N(m_{RO^t}, s_{RO^t}). \tag{67}$$

Then, following the financial principles behind the definition the stochastic oscillator, we consider that large values of either the compound stochastic or compound random oscillator indicate that the stock is overbought, thereby suggesting that it might be a good time to sell the stock. Similarly, small values of these oscillators indicate that the stock is oversold, thereby suggesting that it might be a good time to buy the stock. Therefore, we introduce the variable *act* with possible values buy and sell. The threshold for overbought is set to 65 for both technicals and the threshold for oversold for the both technicals is set to 35. In other words, the technical is low if its value is below 35, it is high if its value is above 65 and it is at a medium level if its value is between 35 and 65. This leads us to the assumption variables a_{CSO^t}, a_{CRO^t}, which both have the same set of possible values, namely low, medium and high. Clearly, the probability of each assumption value of a_{CSO^t} is given by

$$P(a_{CSO^t} = \text{low}) = \int_0^{35} \varphi_{CSO^t}(x)dx,$$
$$P(a_{CSO^t} = \text{medium}) = \int_{35}^{65} \varphi_{CSO^t}(x)dx,$$
$$P(a_{CSO^t} = \text{high}) = \int_{65}^{100} \varphi_{CSO^t}(x)dx,$$

where φ_{CSO^t} is the density function of CSO^t given in Eq. (66). The probability of the three assumption values for the variable CRO^t can be computed in a similar way using the density function given in Eq. (67).

The random variables CSO^t and CRO^t are two features of the stock that could be used to help decide between buying and selling at a given time t. Intuitively, if either of the two features is high, then sell, whereas if either of them is low, then buy (other rules could also be considered, for example if both features are low then buy and if both features are high then sell). This leads to the rules

$$(a_{CSO^t} = \text{low}) \rightarrow (\text{act} = \text{buy}), \quad (a_{CSO^t} = \text{high}) \rightarrow (\text{act} = \text{sell}),$$
$$(a_{CRO^t} = \text{low}) \rightarrow (\text{act} = \text{buy}), \quad (a_{CRO^t} = \text{high}) \rightarrow (\text{act} = \text{sell}).$$

These rules, together with the corresponding assumption probabilities, form a PAS model which can be used to find degrees of support for selling or buying the stock at a given time t. The quasi-support for selling the stock is

$$QS(\text{sell}) = (a_{CSO^t} = \text{high}) \vee (a_{CRO^t} = \text{high})$$

and therefore the degree of quasi-support for selling is

$$qs(\text{sell}) = P(a_{CSO^t} = \text{high}) + P(a_{CRO^t} = \text{high}) - P(a_{CSO^t} = \text{high} \wedge a_{CRO^t} = \text{high}).$$

Now, if the stochastic oscillator and the random oscillator were independent, then the random variables CSO^t and CRO^t would also be independent and we would have

$$P(a_{CSO^t} = \text{high} \wedge a_{CRO^t} = \text{high}) = P(a_{CSO^t} = \text{high}) \cdot P(a_{CRO^t} = \text{high}).$$

However, in reality, the stochastic and random oscillators are not independent, which implies that

$$P(a_{CSO^t} = \text{high} \wedge a_{CRO^t} = \text{high}) = \int_{65}^{100} \int_{65}^{100} \varphi_{CSO^t, CRO^t}(x, y)dx\,dy,$$

where φ_{CSO^t, CRO^t} is the joint density function of the bivariate random variable (CSO^t, CRO^t). We consider that this random variable is normally distributed with mean

$$m^t = (m_{CSO^t}, m_{CRO^t})$$

and correlation coefficient

$$\rho^t = R^t(SO^t, RO^t),$$

where $R^t(SO^t, RO^t)$ is the correlation coefficient of the N data points

$$(SO^k, RO^k), \quad k = t, \ldots, t - N + 1.$$

The degree of contradiction can be found in the same way, which then allows us to compute the degree of support for selling using Eq. (28). Fig. 1 shows the degree of support for selling under the dependent and independent conditions for 51 trading days of the Amazon.com stock when the compounding period is $N = 14$. This figure clearly shows that incorrectly considering that the two features are independent provides results that are considerably higher than those obtained by taking into account the dependency between the two features. Interestingly, if the length of the compounding period N is increased, then a smoothing effect on the degree of support for selling is created, as shown in Figs. 2 and 3 obtained by taking $N = 21$ and $N = 28$, respectively.

4. Feature extraction

In many practical situations, for example in defense applications, an object's features are not directly measured by the sensors. Instead, they measure some characteristics of the object, which we call signatures, and then the features are extracted from these signatures. Signatures are precise and deterministic numerical quantities associated with an object, for example the spectral emissive power of the object.

4.1. Noisy sensor measurements

At any given time, the values of the signatures are not known exactly because of the imprecision of the measurements provided by the sensors. Let us consider a situation where there are two signatures

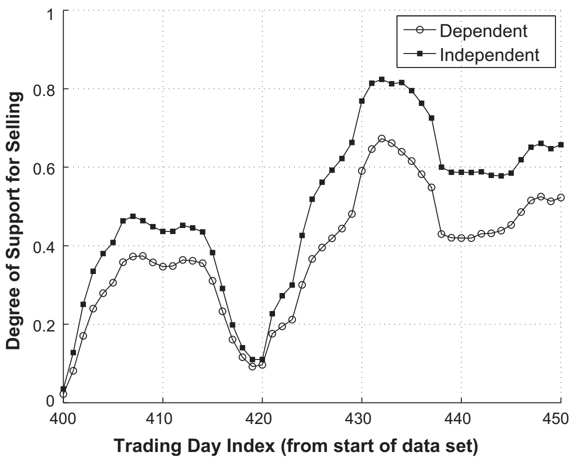

Fig. 1. Degree of support for selling the Amazon.com stock with compounding period $N = 14$.

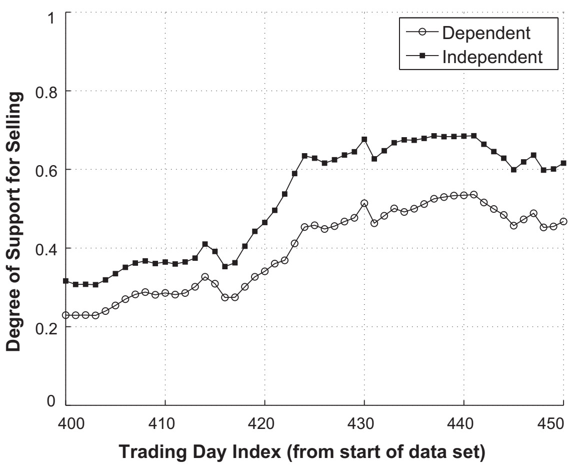

Fig. 2. Degree of support for selling the Amazon.com stock with compounding period $N = 21$.

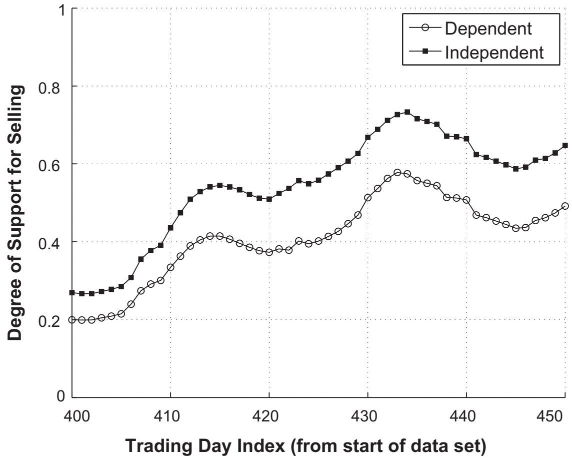

Fig. 3. Degree of support for selling the Amazon.com stock with compounding period $N = 28$.

$$g_1(t), \quad g_2(t),$$

where $g_1(t)$ represents the object's first signature at time t and $g_2(t)$ represents the object's second signature at time t. Since the sensors report the value of these signatures at discrete times $k = 0, 1, 2, \ldots, T$, let

$$S_1(k), \quad S_2(k)$$

denote the measurement of the signatures $g_1(k), g_2(k)$ as reported by the sensors at time k. There is a difference between $g_j(k)$ and $S_j(k)$ because of the imprecision in the measurement process: the sensors do not have the capability of measuring the signatures perfectly, namely, the measurements are random perturbations of the exact signatures. Specifically, we assume that

$$S_j(k) = g_j(k) + w_j(k), \tag{68}$$

where

$$w_j(k) \sim N(0, \delta(k)) \tag{69}$$

is a random noise that is normally distributed with mean 0 and standard deviation $\delta(k)$. To account for the increased signal-to-noise ratio over time, we assume that $\delta(k)$ is a decreasing function of k, so that the measurements become less noisy as time proceeds. For example, over a time horizon of $T = 500$, we can take

$$\delta(k) = \left(2 - \frac{1.4}{500}k\right)^2. \tag{70}$$

The graph of $\delta(k)$ is shown in Fig. 4.

In addition, we assume that the noises $w_1(k)$ and $w_2(k)$ are stochastically independent because the sensors that measure the two signatures are different.

Let us now explain how the features are extracted from the signatures. First, let $f_1(k)$ and $f_2(k)$ denote the two features at time k. These features are extracted from the two signatures at time $l = 0, 1, \ldots, k$. In other words, for the first feature, it is assumed that there is a function

$$H_1 : \mathbb{R}^{k+1} \times \mathbb{R}^{k+1} \rightarrow \mathbb{R}$$

such that

$$f_1(k) = H_1(g_1(0), \ldots, g_1(k); g_2(0), \ldots, g_2(k)).$$

Similarly, for the second feature, it is assumed that there is a function

$$H_2 : \mathbb{R}^{k+1} \times \mathbb{R}^{k+1} \rightarrow \mathbb{R}$$

such that

$$f_2(k) = H_2(g_1(0), \ldots, g_1(k); g_2(0), \ldots, g_2(k)).$$

Therefore, if we define the function

$$H : \mathbb{R}^{k+1} \times \mathbb{R}^{k+1} \rightarrow \mathbb{R}^2$$

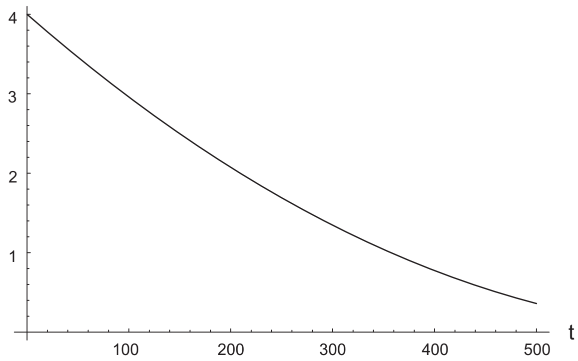

Fig. 4. Quadratic decay of the noise's standard deviation over time.

by

$$H(u;v) = \begin{pmatrix} H_1(u;v) \\ H_2(u;v) \end{pmatrix},$$

then

$$\begin{pmatrix} f_1(k) \\ f_2(k) \end{pmatrix} = H(g_1(0),\ldots,g_1(k);g_2(0),\ldots,g_2(k)). \tag{71}$$

The two features f_1 and f_2 at time k are obtained by applying the function H to the object's signatures

$$(g_1(0),\ldots,g_1(k);g_2(0),\ldots,g_2(k)).$$

This shows that the features are extracted from the signatures by means of the function H, which is presumed to be known, as is often the case in practice.

4.2. Fiducial distributions

When extracting the features from the signatures, we are confronted with the difficulty that the signatures $g_j(l)$ are not known exactly, only their noisy measurements $S_j(l)$ are known. But we can use Eq. (68) to derive information about the unknown signature value $g_j(l)$. It is important to note that only the sensor measurement $S_j(l)$ is known. Then, using (68), we can derive the so-called fiducial distribution of $g_j(l)$. It is the normal distribution with mean $S_j(l)$ and standard deviation $\delta(l)$. This fiducial distribution represents our knowledge about the fixed but unknown quantity $g_j(l)$. This is not the same as saying that $g_j(l)$ is a random variable having this distribution. The fiducial distribution allows us to compute the degree of support for the hypothesis stating that the signature $g_j(l)$ is within a specified interval by integrating it over that interval. In our case, the fiducial distribution of $g_j(l)$ is a probability distribution, namely the distribution $N(S_j(l),\delta(l))$, which is completely known as soon as the sensor measurement $S_j(l)$ is reported. By abuse of notation, since $g_j(l)$ is not a random variable, we write

$$g_j(l) \sim N(S_j(l),\delta(l)). \tag{72}$$

Again, the fiducial distribution is completely known because $S_j(l)$ is the measurement reported by the sensor and $\delta(l)$ is computed using Eq. (70). R.A. Fisher first introduced the notion of fiducial distribution when he was looking for ways to reason probabilistically from sample data to posterior distributions without the need of prior distributions as in Bayesian statistics [31]. A modern presentation of some aspects of the fiducial theory can be found in [32,33].

According to Eq. (71), the pair of features at time k is a function of the signatures $g_j(l)$, whose fiducial distribution is given in (72). Therefore, the joint fiducial distribution of the features at time k is obtained by applying the function H to a collection of known normal distributions, namely the fiducial distributions of the signatures $g_j(l)$. If the function H is invertible and differentiable, for example if H is linear, then we can use the classical transformation theorem to determine the joint fiducial distribution of $(f_1(k),f_2(k))$. However, in practice, the function H is most likely complicated and not invertible, thereby preventing us from applying the transformation theorem.

4.3. Kernel density estimation

In this section we present a method for finding the joint fiducial distribution of the two features at time k when the transformation theorem is not applicable, for example when H is not invertible. The first step is to derive a scatter plot of the features vector $(f_1(k),f_2(k))$. The scatter plot consists of n data points

$$(f_{i1},f_{i2}), \quad i = 1,\ldots n \tag{73}$$

that can be grouped into the data matrix

$$F = \begin{pmatrix} f_{11} & f_{12} \\ \vdots & \vdots \\ f_{i1} & f_{i2} \\ \vdots & \vdots \\ f_{n1} & f_{n2} \end{pmatrix}. \tag{74}$$

This data matrix F is obtained by applying the following procedure:
For $i = 1$ to n, do the following:

(1) generate a vector

$$x = (x_0, x_1, \ldots, x_l, \ldots, x_k),$$

where x_l is a random realization of the distribution $N(S_1(l), \delta(l))$.

(2) generate a vector

$$y = (y_0, y_1, \ldots, y_l, \ldots, y_k),$$

where y_l is a random realization of the distribution $N(S_2(l), \delta(l))$.

(3) compute the data point

$$(f_{i1}, f_{i2}) = H(x; y),$$

i.e.

$$f_{i1} = H_1(x; y),$$
$$f_{i2} = H_2(x; y).$$

Once the data matrix F is computed, in the second step, we determine the joint fiducial distribution of $(f_1(k), f_2(k))$ by applying Gaussian kernel density estimation to the data matrix F [34]. The result is the joint density function

$$f(x_1, x_2) = n^{-\frac{2}{3}}(\widehat{\sigma}_1 \widehat{\sigma}_2)^{-1} \cdot \sum_{i=1}^{n} \varphi\left(\frac{x_1 - f_{i1}}{h_1}\right) \cdot \varphi\left(\frac{x_2 - f_{i2}}{h_2}\right), \tag{75}$$

where

(1) φ is the density function of the normal distribution with mean 0 and standard deviation 1, i.e.

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

(2) $\widehat{\sigma}_1$ is the standard deviation of the f_1 data points, i.e.

$$\widehat{\sigma}_1 = \left(\frac{1}{n-1} \sum_{i=1}^{n} (f_{i1} - \bar{f}_1)^2\right)^{0.5}$$

(3) $\widehat{\sigma}_2$ is the standard deviation of the f_2 data points, i.e.

$$\widehat{\sigma}_2 = \left(\frac{1}{n-1} \sum_{i=1}^{n} (f_{i2} - \bar{f}_2)^2\right)^{0.5}$$

(4) h_1 is the binwidth for f_1, which is given by

$$h_1 = n^{-\frac{1}{6}} \cdot \widehat{\sigma}_1$$

(5) h_2 is the binwidth for f_2, which is given by.

$$h_2 = n^{-\frac{1}{6}} \cdot \widehat{\sigma}_2.$$

This density function $f(x_1, x_2)$ allows us to compute the probability that the two features are within a rectangle

$$I = I_1 \times I_2$$

with

$$I_1 = [a, b], \quad I_2 = [c, d].$$

It can be shown that

$$P(I_1 \wedge I_2) = \frac{1}{n} \cdot \sum_{i=1}^{n} (F_{i1}(b) - F_{i1}(a)) \cdot (F_{i2}(d) - F_{i2}(c)), \tag{76}$$

where F_{i1} is the cumulative distribution function of the normal distribution with mean f_{i1} and standard deviation h_1 and F_{i2} is the cumulative distribution function of the normal distribution with mean f_{i2} and standard deviation h_2.

The knowledge of the joint density $f(x_1, x_2)$ can also be used to compute the marginal distribution for both features f_1 and f_2. The marginal distribution of f_1 is the function

$$\hat{f}_1 : \mathbb{R} \to [0, \infty)$$

defined by

$$\hat{f}_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2$$

and it can be proved that

$$\hat{f}_1(x_1) = n^{-\frac{2}{3}} (\hat{\sigma}_1 \hat{\sigma}_2)^{-1} h_2 \sum_{i=1}^{n} \varphi\left(\frac{x_1 - f_{i1}}{h_1}\right). \tag{77}$$

Similarly, the marginal distribution of the second feature f_2 is the function

$$\hat{f}_2 : \mathbb{R} \to [0, \infty)$$

defined by

$$\hat{f}_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1$$

and it can be proved that

$$\hat{f}_2(x_2) = n^{-\frac{2}{3}} (\hat{\sigma}_1 \hat{\sigma}_2)^{-1} h_1 \sum_{i=1}^{n} \varphi\left(\frac{x_2 - f_{i2}}{h_2}\right). \tag{78}$$

These marginal density functions can be used to compute the probability that a feature lies within a certain interval. In addition, if the features were assumed independent, the joint density would be

$$f_{ind}(x_1, x_2) = \hat{f}_1(x_1) \cdot \hat{f}_2(x_2)$$

and it can be proved that

$$f_{ind}(x_1, x_2) = n^{-\frac{5}{3}} (\hat{\sigma}_1 \hat{\sigma}_2)^{-1} \cdot \left(\sum_{i=1}^{n} \varphi\left(\frac{x_1 - f_{i1}}{h_1}\right) \right) \cdot \left(\sum_{i=1}^{n} \varphi\left(\frac{x_2 - f_{i2}}{h_2}\right) \right).$$

Then we can show that the probability of a rectangle would be

$$P_{ind}(I_1 \wedge I_2) = \frac{1}{n^2} \cdot \left(\sum_{i=1}^{n} (F_{i1}(b) - F_{i1}(a)) \right) \cdot \left(\sum_{i=1}^{n} (F_{i2}(d) - F_{i2}(c)) \right).$$

4.4. Numerical example

In this section we present a numerical illustration of the method described above. Although in reality they are not known, for the sake of the simulation, we suppose that the first signature is

$$g_1(t) = 30 \exp(-0.0085t) + 2 \sin(2\pi 0.015t + 0.785) + \sin(2\pi 0.09t + 1.965) + 6 \tag{79}$$

and the second signature is

$$g_2(t) = 50 \exp(-0.0085t) + 1.5 \sin(2\pi 0.015t + 2.745) + 3 \sin(2\pi 0.09t + 2.925) + 20 \tag{80}$$

The graphs of these signatures are given in Fig. 5.

These signatures are used to simulate sensor measurements $S_j(k), k = 0, \ldots, 500$ according to Eq. (68), namely

$$S_j(k) = g_j(k) + w_j(k).$$

The graphs of the sensor measurements that are obtained for both signatures are given in Fig. 6.

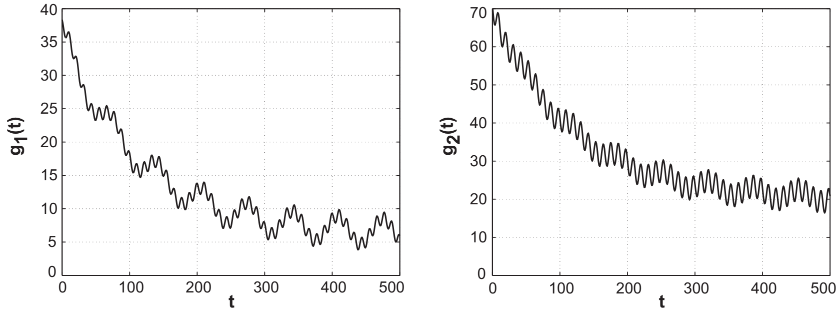

Fig. 5. The graphs of the two signatures.

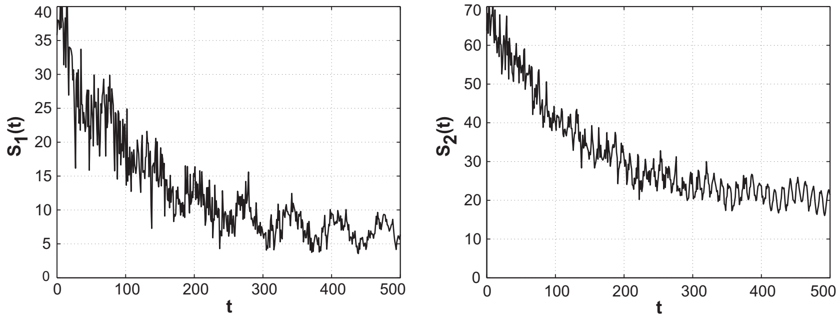

Fig. 6. The sensor measurements.

Suppose we are now at time $t = 5$. For our illustration, we need to define the functions that are used to extract the two features from the two signatures. In other words, we need to define the function

$$H_1 : \mathbb{R}^6 \times \mathbb{R}^6 \rightarrow \mathbb{R},$$

where

$$H_1(u_0, \ldots, u_5; v_0, \ldots, v_5)$$

is the value of the first feature at time 5 when $(u_0, \ldots u_5)$ is the vector of the first signature and $(v_0, \ldots, v_5)$ is the vector of the second signature. In this illustration, we take

$$H_1(u_0, \ldots, u_5; v_0, \ldots, v_5) = z,$$

where

$$z = Max\ \{Abs(DFT(u_0, \ldots, u_5))^* \cdot Abs(DFT(v_0, \ldots, v_5))^*\}$$

and *DFT* is the discrete Fourier transform, *Abs* is the modulus and * means that the first component of the vector is discarded. Since

$$Abs(z_1 \cdot z_2) = Abs(z_1) \cdot Abs(z_2)$$

for any complex numbers z_1 and z_2, we also have

$$z = Max\ \{Abs(DFT(u_0, \ldots, u_5) \cdot DFT(v_0, \ldots, v_5))^*\}.$$

Similarly, for the extraction of the second feature, we need to define the function

$$H_2 : \mathbb{R}^6 \times \mathbb{R}^6 \rightarrow \mathbb{R},$$

where

$$H_2(u_0, \ldots, u_5; v_0, \ldots, v_5)$$

represents the value of the second feature at time 5 if $(u_0, \ldots u_5)$ is the vector of the first signature and $(v_0, \ldots, v_5)$ is the vector of the second signature. In this example, let us take

$$H_2(u_0, \ldots, u_5; v_0, \ldots, v_5) = u_5^2 + v_5^2.$$

This means that the value of the second feature at time 5 only depends on the object's two signatures u_5 and v_5 at the current time 5. The definition of the functions H_1 and H_2 completely specifies the function

$$H : \mathbb{R}^6 \times \mathbb{R}^6 \rightarrow \mathbb{R}^2 \tag{81}$$

representing the feature extraction mechanism. It can easily be shown that the exact values of the two features at time 5 is

$$(f_1, f_2) = (26.913,\ 5665.653). \tag{82}$$

In order to determine the joint fiducial density of the two features at time 5, the next step is to create the features' data matrix F that is needed for the application of the kernel density estimation technique. We generate $n = 2000$ data points (f_{i1}, f_{i2}), which are shown in the scatter plot displayed in Fig. 7.

Based on these points, the joint density function of the features at time 5 is obtained with Eq. (75). Fig. 8 shows the density plot of the joint density function $f(x_1, x_2)$, where lighter areas represent larger values.

This figure gives us an idea of the dependence between the two features, namely a weak negative relationship between the two features. The marginal density functions of the features are shown in Fig. 9.

Let us now use these two features f_1 and f_2 to classify an object into one of two categories c_1 or c_2. The knowledge base consists of the probabilistic argumentation system containing the rules

$$\begin{aligned} I_1 \wedge a_1 &\rightarrow c_1, \\ I_2 \wedge a_2 &\rightarrow c_1, \\ J_1 \wedge a_3 &\rightarrow c_2, \\ J_2 \wedge a_4 &\rightarrow c_2, \end{aligned}$$

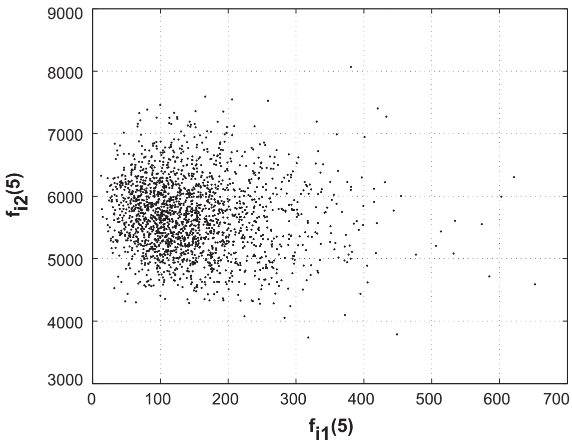

Fig. 7. Scatter plot of the data points (f_{i1}, f_{i2}).

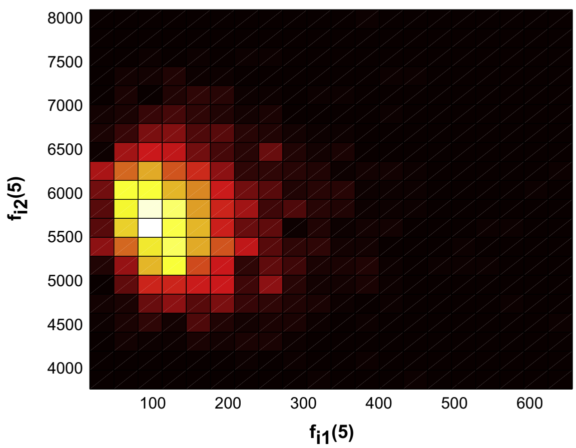

Fig. 8. Joint density function of the two features.

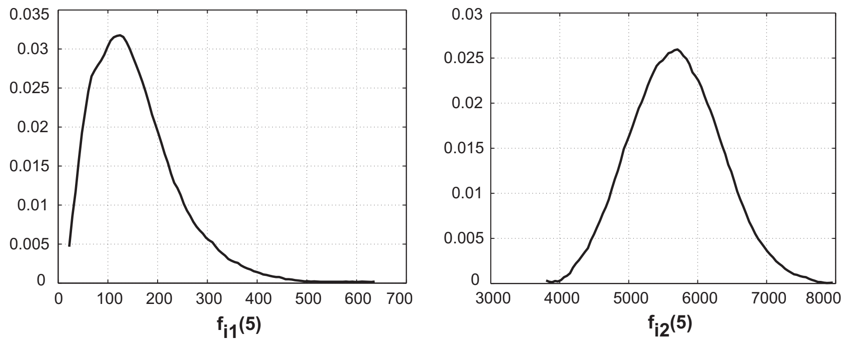

Fig. 9. The marginal distributions of the two features.

where $I_1 = [10, 20]$ and $J_1 = [30, 90]$ are two intervals for the first feature and $I_2 = [3500, 5000]$ and $J_2 = [5500, 7000]$ are two intervals for the second feature. Clearly, the quasi-support of c_1 is

$$QS(c_1) = (I_1 \wedge a_1) \vee (I_2 \wedge a_2).$$

Therefore, the degree of quasi-support of c_1 is

$$qs(c_1) = P(I_1 \wedge a_1) + P(I_2 \wedge a_2) - P(I_1 \wedge I_2 \wedge a_1 \wedge a_2).$$

If p_i denotes the probability of a_i, then this degree of quasi-support can be written as

$$qs(c_1) = p_1P(I_1) + p_2P(I_2) - p_1p_2P(I_1 \wedge I_2).$$

Similarly, the quasi-support of the contradiction is

$$QS(\perp) = (a_1 \wedge a_4 \wedge I_1 \wedge J_2) \vee (a_2 \wedge a_3 \wedge I_2 \wedge J_1)$$

and therefore the degree of contradiction is

$$qs(\perp) = P(QS(\perp)) = P(a_1 \wedge a_4 \wedge I_1 \wedge J_2) + P(a_2 \wedge a_3 \wedge I_2 \wedge J_1)$$

since the two conjunctive terms in $QS(\perp)$ are incompatible. This shows that the degree of contradiction is

$$qs(\perp) = p_1p_4P(I_1 \wedge J_2) + p_2p_3P(I_2 \wedge J_1).$$

Using the marginal density functions of the features we find

$$P(I_1) = 0.172, \quad P(I_2) = 0.324, \quad P(J_1) = 0.514, \quad P(J_2) = 0.332$$

and using the joint density function of the two features we find

$$P(I_1 \wedge I_2) = 0.033, \quad P(I_1 \wedge J_2) = 0.077, \quad P(J_1 \wedge I_2) = 0.213, \quad P(J_1 \wedge J_2) = 0.132.$$

If we take $p_i = 0.9$ for $i = 1, \ldots, 4$, then the degree of support of the hypothesis c_1 is

$$sp(c_1) = \frac{qs(c_1) - qs(\perp)}{1 - qs(\perp)} = 0.241.$$

Similarly, the quasi-support of c_2 is

$$QS(c_2) = (J_1 \wedge a_3) \vee (J_2 \wedge a_4)$$

and the degree of quasi-support of c_2 is

$$qs(c_2) = p_3P(J_1) + p_4P(J_2) - p_3p_4P(J_1 \wedge J_2)$$

and hence the degree of support of c_2 is

$$sp(c_2) = \frac{qs(c_2) - qs(\perp)}{1 - qs(\perp)} = 0.548.$$

Since $sp(\emptyset) = 0$ and $sp(c_1, c_2) = 1$ the support function sp on $\{c_1, c_2\}$ is completely known and its pignistic probabilities are

$$BetP(c_1) = 0.3495, \quad BelP(c_2) = 0.6505.$$

Given the information provided by the sensors and the two dependent features, we conclude that the object must be placed into category c_2. It is interesting to note that our method also works in case of independence of the two features, which means that we don't need to worry ahead of time whether or not the two features are dependent.

5. Conclusion

We have presented a model and corresponding techniques to deal with situations where the features are both noisy and dependent. The model is based on the general tools and concepts provided by probabilistic argumentation systems, which makes it easier for the knowledge engineer to avoid including items of information that are not fully warranted. For example, the model does not require conditional density distributions of features given each possible class for the object, unlike Bayesian analysis or the method proposed by Ristic and Smets [7]. Because of the soundness of the probabilistic argumentation system framework, it is our belief that the techniques presented in this paper can be successfully applied to a variety of classification problems, in particular automatic target recognition.

Acknowledgements

The authors would like to thank the reviewers of the paper for their helpful comments.

References

[1] T. Denoeux, A k-nearest neighbor classification rule based on Dempster–Shafer theory, IEEE Transactions on Systems, Man and Cybernetics 25 (05) (1995) 804–813.
[2] T. Denoeux, A neural network classifier based on Dempster–Shafer theory, IEEE Transactions on Systems, Man and Cybernetics A 30 (2) (2000) 131–150.
[3] T. Denoeux, L.M. Zouhal, Handling possibilistic labels in pattern classification using evidential reasoning, Fuzzy Sets and Systems 122 (3) (2001) 47–62.
[4] T. Denoeux, P. Smets, Classification using belief functions: the relationship between the case-based and model-based approaches, IEEE Transactions on Systems, Man and Cybernetics B 36 (6) (2006).
[5] B. Quost, T. Denoeux, M.-H. Masson, Pairwise classifier combination using belief functions, Pattern Recognition Letters 28 (5) (2007) 644–653.
[6] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers, Machine Learning 29 (2-3) (1997) 131–163.
[7] B. Ristic, P. Smets, Belief function theory on the continuous space with an application to model based classification, in: Proceedings of the Conference Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2004, Perugia, Italy, 2004, pp. 1119–1126.
[8] P. Smets, Decision making in the TBM: the necessity of the pignistic transformation, International Journal of Approximate Reasoning (38) (2005) 133–147.
[9] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
[10] J. Kohlas, P.A. Monney, A Mathematical Theory of Hints, Lecture Notes in Economics and Mathematical Systems, vol. 425, Springer, Berlin, 1995.
[11] P. Smets, R. Kennes, The transferable belief model, Artificial Intelligence 66 (1994) 191–243.
[12] P. Smets, Practical uses of belief functions, in: K.B. Laskey, H. Prade (Eds.), Uncertainty in Artificial Intelligence, vol. 15, Stockholm, Sweden, 1999.
[13] P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, International Journal of Approximate Reasoning 9 (1993) 1–35.
[14] P. Smets, The transferable model for quantified belief representation, in: P. Smets (Ed.), Quantified Representation of Uncertainty and Imprecision, Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 1, Kluwer, Dordrech, 1998, pp. 267–301.
[15] A.P. Dempster, A generalization of Bayesian inference, Journal of the Royal Statistical Society Series B 30 (1968) 205–247.
[16] A. Aregui, T. Denoeux, Constructing consonant belief functions from sample data using confidence sets of pignistic probabilities, International Journal of Approximate Reasoning 49 (3) (2008) 575–594.
[17] R. Haenni, J. Kohlas, N. Lehmann, Probabilistic argumentation systems, in: J. Kohlas, S. Moral (Eds.), Algorithms for Uncertainty and Defeasible Reasoning, Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 5, Kluwer, Dordrecht, 2000.
[18] J. Kohlas, R. Haenni, et al, Propositional information systems, Journal of Logic and Computation 9 (5) (1999) 651–681.
[19] T. Baker, M. Chan, T. Hansen, P.-A. Monney, A Java implementation of the probabilistic argumentation system for data fusion in missile defense applications, in: Proceedings of SPIE, Defense and Security Symposium on Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Application, vol. 5434, Orlando, FL, USA, April 12–16, 2004, pp. 176–186.
[20] R. Haenni, Probabilistic argumentation, Journal of Applied Logic 7 (2) (2009) 155–176.
[21] K.B. Laskey, P.E. Lehner, Assumptions beliefs and probabilities, Artificial Intelligence 41 (1989) 65–77.
[22] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufman, 1988.
[23] J. Kohlas, Information Algebras. Generic Structures for Inference. Discrete Mathematics and Theoretical Computer Science, Springer, London, 2003.
[24] J. Kohlas, P.-A. Monney, Statistical Information. Assumption-Based Statistical Inference. Heldermann Verlag, Sigma Series in Stochastics, vol. 3, ISBN 978-3-88538-303-1, Lemgo, Germany, 2008.
[25] P.A. Monney, A Mathematical Theory of Arguments for Statistical Evidence. Contributions to Statistics, Physica Verlag (Springer), 2003.
[26] J. Kohlas, Uncertain information: random variables in graded semilattices, International Journal of Approximate Reasoning 46 (1) (2007) 17–34.
[27] J. Kohlas, P.A. Monney, An algebraic theory for statistical information based on the theory of hints, International Journal of Approximate Reasoning 48 (2) (2008) 378–398.
[28] R. Haenni, N. Lehmann, ABEL: an interactive tool for probabilistic argumentative reasoning, in: ECSQARU'03, Seventh European Conference on Symbolic and Quantitative Approaches to Reasoning under Uncertainty, Aalborg, Denmark, 2003, pp. 588–593.
[29] P.A. Monney, R. Bertschy, A generalization of the algorithm of Heidtmann to non-monotone formulas, Journal of Computational and Applied Mathematics 76 (1996) 55–76.
[30] P.A. Monney, M. Chan, Modeling dependence in Dempster–Shafer theory, International Journal of Uncertainty Fuzziness and Knowledge-Based Systems 15 (1) (2007) 93–113.
[31] R.A. Fisher, The fiducial argument in statistical inference, Annals of Eugenics 9 (1935) 391–398.
[32] A.P. David, M. Stone, The functional-model basis of fiducial inference, The Annals of Statistics 10 (4) (1982) 1054–1067.
[33] F. Hampel, The proper fiducial argument, Technical Report 114, Swiss Federal Institute of Technology (ETH), CH-8092 Zurich, Switzerland, 2003.
[34] B. Silverman, Density Estimation for Statistics and Data Analysis, Chapman & Hall, London, 1986.